A common protein fold and similar active site in two distinct families of β-glycanases

Roberto Dominguez[1], Hélène Souchon[1], Silvia Spinelli[1,4], Zbigniew Dauter[3], Keith S. Wilson[3], Sylvie Chauvaux[2], Pierre Béguin[2] and Pedro M. Alzari[1]

The structure of *Clostridium thermocellum* endoglucanase CelC, a member of the largest cellulase family (family A), has been determined at 2.15 Å resolution. The protein folds into an $(\alpha/\beta)_8$ barrel, with a deep active-site cleft generated by the insertion of a helical subdomain. The structure of the catalytic core of xylanase XynZ, which belongs to xylanase family F, has been determined at 1.4 Å resolution. In spite of significant differences in substrate specificity and structure (including the absence of the helical subdomain), the general polypeptide folding pattern, architecture of the active site and catalytic mechanism of XynZ and CelC are similar, suggesting a common evolutionary origin.

[1]Unité d'Immunologie Structurale, URA 1961 CNRS, Institut Pasteur, 75724 Paris Cedex 15, France
[2]Unité de Physiologie Cellulaire, URA 1300 CNRS, Institut Pasteur, 75724 Paris Cedex 15, France
[3]EMBL Hamburg Outstation, c/o DESY, Notkestrasse 85, D-2000 Hamburg 52, Germany

[4]Present address: LCCMB, Faculté de Médecine, Secteur Nord, 13916 Marseille Cedex 20, France

Correspondence should be addressed to P.M.A.

Cellulases and hemicellulases offer an excellent opportunity to study structure-function relationships in a wide range of enzymes having closely related catalytic activities but a variety of different structures. Sequence analysis and biochemical studies have shown that many of these enzymes contain a catalytic domain as well as other polypeptide subunits which are usually not directly involved in catalysis (cellulose-binding domains, for example)[1]. Based on sequence comparisons, the catalytic domains of cellulolytic and xylanolytic enzymes can be grouped into several families. In spite of sometimes very low identity scores, enzymes belonging to the same family have been predicted to share a similar polypeptide folding pattern and a similar active-site architecture including conserved catalytic residues[2,3]. Available evidence indicates that β-glycanases of the same family proceed by the same type of mechanism, leading either to inversion or to retention of configuration at the anomeric carbon. By contrast, enzymes belonging to different families may have different mechanisms, and quite different protein folds[4]. It is therefore of interest to determine the three-dimensional structure and mechanism of β-glycanases belonging to different families, in order to provide patterns after which structurally related enzymes can be modelled. This study presents and compares the structures of endoglucanase CelC (Fig. 1) and xylanase XynZ of *Clostridium thermocellum* (Figs 2, 3). CelC is a member of family A cellulases, for which more than 65 sequences (but no previous structural information) are available. XynZ is a member of family F xylanases, for which at least 27 sequences are known.

Structure of endoglucanase CelC

The structure of *C. thermocellum* endoglucanase CelC, a member of cellulase family A, was determined by X-ray crystallography at 2.15 Å resolution using multiple isomorphous replacement and density averaging between two crystal forms. The protein is a cylindrical $(\alpha/\beta)_8$ barrel (Fig. 1) with two β-bulges at strands 3 and 7 and an acidic cleft containing the active site on the carboxy-terminal side of the barrel. First observed in triose phosphate isomerase[5], the $(\alpha/\beta)_8$ barrel topology was subsequently found in enzymes with a wide variety of unrelated activities, including various glycohydrolases such as β-galactosidases, neuraminidases, α- and β-amylases, lichenases, xylanase, and chitinases.

A segment of 54 amino acids adjacent to the active-site cleft of CelC folds into a distinct subdomain consisting of four α-helices and a short two-stranded β-structure. Inserted between strand 6 and helix 6 of the α/β barrel, this subdomain extends the top of the barrel on one side, thus creating a deep substrate-binding cleft. This structural element is present in some, but not all enzymes of family A. Several endoglucanases of subfamily A2, for example, carry a deletion of exactly 54 residues in this region with respect to CelC[6].

The active site and hydrolytic mechanism

Substrate hydrolysis by CelC proceeds by a two-step acid-base mechanism leading to retention of configuration at the anomeric carbon[4]. Glu 280, identified as the nucleophilic residue involved in catalysis[6], lies at the bottom of the active-site crevice. The carboxylate group of Glu 280, partially stacked between conserved residues

Trp 313 and His 198, is within hydrogen-bonding distance of Tyr 200 (2.8 Å) and Arg 46 (3.0 Å;Fig. 5*a*).

The structure of a complex between an inactive mutant of CelC (Glu 280 → Cys), $CelC_{F280C}$, and cellotriose (which has been determined at 3 Å resolution) shows the substrate bound to the bottom of the crevice (Fig. 4*c*), adjacent to the residue at position 280 and in contact with Asn 139, His 90 and Glu 140. Amino acid substitutions at these three positions strongly inactivate CelC[7] and other homologous cellulases[8,9], and suggest that Glu 140 is the proton donor in the catalytic reaction[7]. The crystal structure of the complex confirms this hypothesis; the glutamate side chain is immersed in a hydrophobic environment, favourably positioned to interact directly with the substrate (Fig. 4*c*). In the unliganded orthorhombic structure, however, Glu 140 projects outwards from the substrate-binding site (Fig.1*a*). The different orientations of Glu 140 are coupled to the conformation of the loop residues 95–103, which presents a well-defined structure in the substrate-bound crystal form (Fig. 4*c*) but is disordered in the unliganded structure. In particular, Phe 97 closes in on the carboxylate group of Glu 140 when ligand is bound. Similar conformational changes associated with substrate binding[10] or pH variation[11] were observed in other β-glycanases. Crystal packing differences might explain, at least in part, the observed structural changes between the unliganded (orthorhombic) and substrate-bound (tetragonal) forms of CelC, but the concerted movement of the two active site loops may also reflect an induced fit effect associated with ligand binding. Additional studies of uncomplexed and complexed CelC in the same crystal form are required to further clarify this issue.

Structure of xylanase XynZ

The structure of the catalytic domain of *C. thermocellum* xylanase XynZ, which belongs to β-glycanase family F, was determined at 1.4 Å resolution. The domain folds into an eight-stranded α/β barrel with an additional amino-terminal α-helix (Fig. 2), a structure similar to that of homologous xylanases from other species[12–14]. The active site is formed by an acidic cleft on the carboxy-terminal side of the β-strands. Amino acid residues in the β-barrel form a continuous hydrogen-bonded network around the cylinder, partially disrupted at strands 3, 4 and 7 by β-bulges containing key functional residues (His 596, Glu 645 and Glu 754). The structure resembles that of CelC in overall shape and orientation of the substrate-binding groove. The similarity extends to specific features such as the two β-bulges at strands 3 and 7, the conformation of some loops at the bottom of the barrel, and the presence of several intramolecular salt bridges. Equivalent regions of the two structures can be superimposed with a root-mean-square deviation of 1.9 Å for 159 corresponding Cα positions (Fig. 3). There are 25 sequence identities between CelC and XynZ; 21 of these occur in structurally equivalent regions either at various framework positions or at key functional positions in the active site. The resulting alignment reveals a better match towards the C-terminal end of helices and N-terminal end of strands (Fig. 3b), consistent with structural changes being required to evolve a distinct specificity.

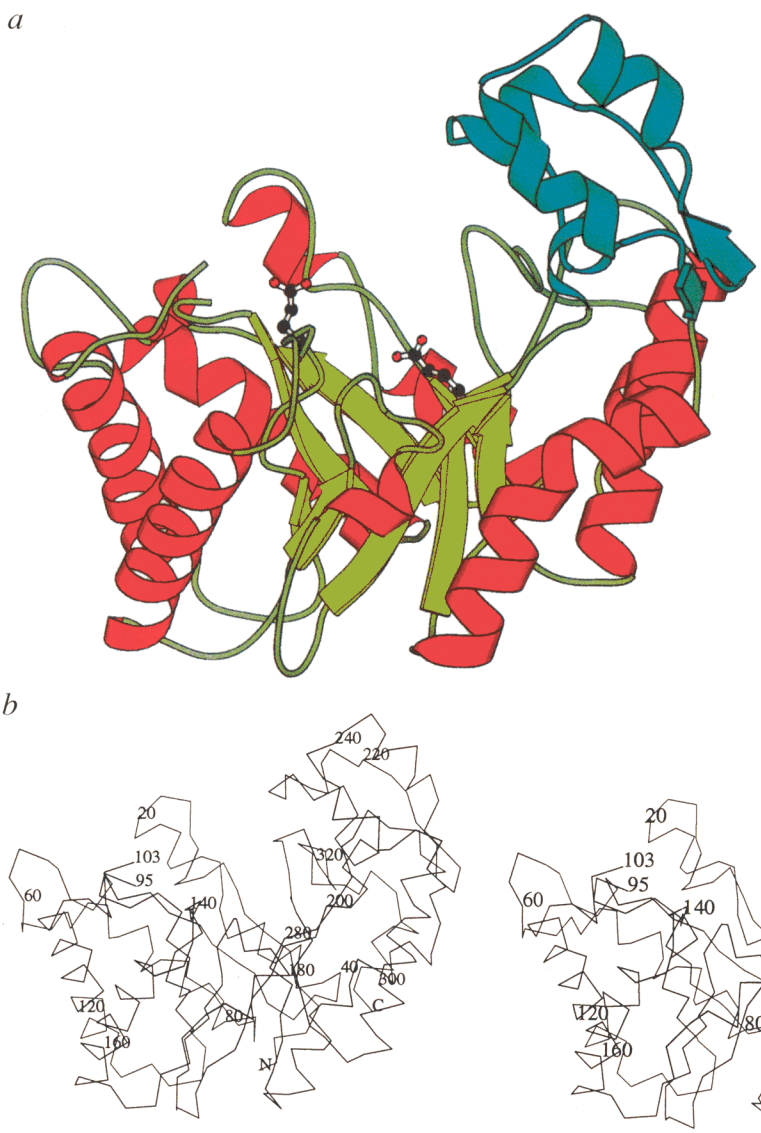

Fig. 1 Schematic view of endoglucanase CelC. *a*, Ribbon diagram (drawn with MOLSCRIPT[39]) showing the overall topology. Helices and strands in the $(\alpha/\beta)_8$ barrel are coloured in red and yellow, respectively. The subdomain inserted between strand 6 and helix 6 is shown in blue. The catalytic residues are shown in ball-and-stick representation. *b*, Cα stereodiagram oriented as in (*a*). Residues 95–103 (not shown) are disordered in the orthorhombic crystal structure.

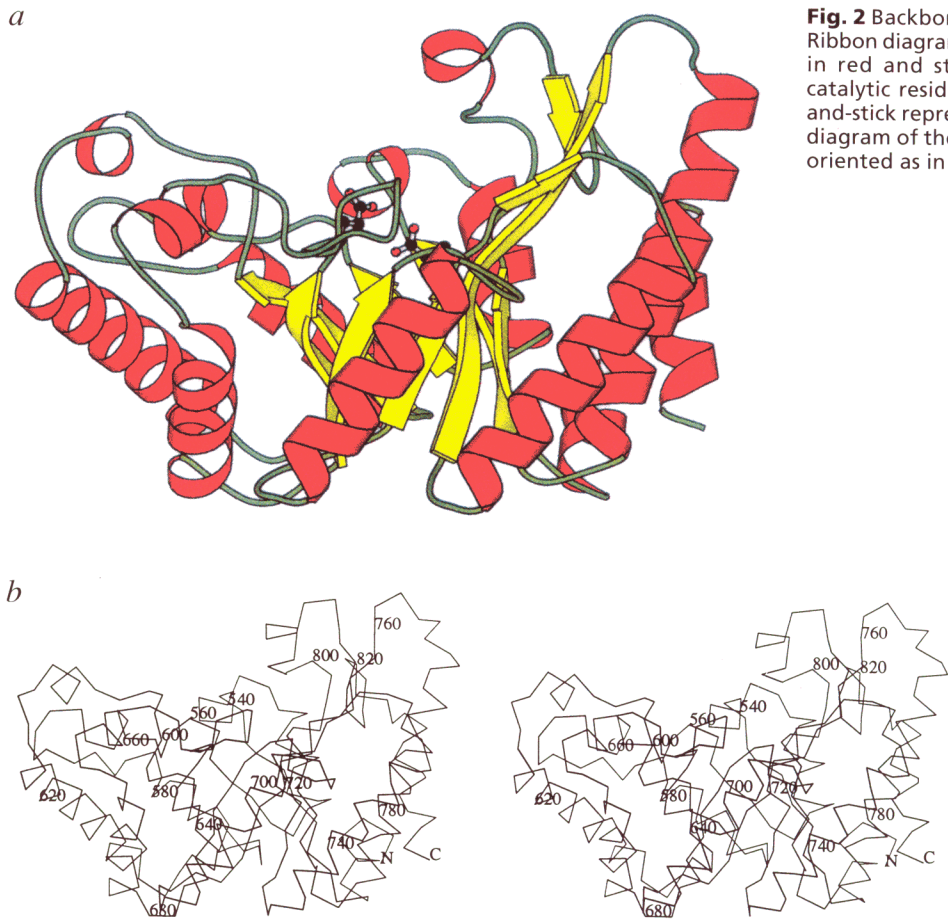

Fig. 2 Backbone structure of XynZ. *a*, Ribbon diagram with helices coloured in red and strands in yellow. The catalytic residues are shown in ball-and-stick representation. *b*, Cα stereo diagram of the molecule. The view is oriented as in Fig. 1.

Xylanase active site

Hydrolysis by XynZ proceeds with retention of an anomeric configuration[4] through a double displacement reaction mechanism. Residue Glu 754, strictly conserved and identified as the nucleophile in other xylanases[15], is positioned at the bottom of an acidic crevice (Fig. 4*b*). Glu 754 is within hydrogen-bonding distance of His 723 (2.80 Å), Asn 687 (3.07 Å) and two water oxygens (2.79, 2.96 Å) which fill a cavity deep inside the active site. The nucleophile is stacked between two conserved residues, Trp 795 and Gln 721 (Fig. 5*b*). More exposed to the solvent, the acid catalyst Glu 645 is positioned in a less restrained stereochemical environment, forming weak hydrogen bonds with Trp 600 and Gln 721. Substrate binding could possibly modify the local structure and charge density around the side chain of Glu 645, thus favouring the protonated state required for catalysis[14].

Similar active sites in different families

The geometry and mechanism of the catalytic centres of CelC and XynZ are strikingly similar. Both enzymes hydrolyze the β-1,4-glycosidic bond through a double displacement mechanism[4]. The catalytic glutamate residues, at equivalent positions in the sequence (Fig. 3*b*), are situated in a similar chemical environment. The nucleophilic carboxylate is stacked between a strictly conserved Trp residue and a polar planar group - histidine in family A cellulases or glutamine in family F xylanases (Fig. 5). Moreover, the protonation state of this glutamate is conditioned by hydrogen-bond interactions with a conserved tyrosine (family A) or histidine (family F) side chain. The proton donor Glu 645 has the same conformation as Glu 140 in the liganded form of CelC. Other equivalent positions include an invariant acidic residue (Glu 136 in CelC, Asp 641 in XynZ) which forms a buried salt bridge in both structures, and an asparagine residue immediately preceding the proton donor glutamate at the end of strand 4 (Fig. 3*b*). Substitution of Asn 139 by alanine or aspartic acid in CelC strongly inactivates the enzyme[7]. Indeed, the asparagine side chain occupies a central position in the active site of the two structures (Fig. 5), close to (but not in contact with) the two catalytic glutamates.

The similar active site architecture of family A cellulases and family F xylanases is also observed in other β-glycanases. The spatial arrangement of active site residues in two barley glycanases with different specificities[16] closely resembles that of CelC and XynZ (Fig. 6). For example, the nucleophile Glu 232 is also stacked between an aromatic and a polar planar group, with its carboxylate group within hydrogen-bonding distance of a tyrosine residue. At equivalent positions in the sequence, Arg 31 fulfills a similar structural role as Arg 46 in CelC, and a bulge at the end of β-strand 4 brings the two polar

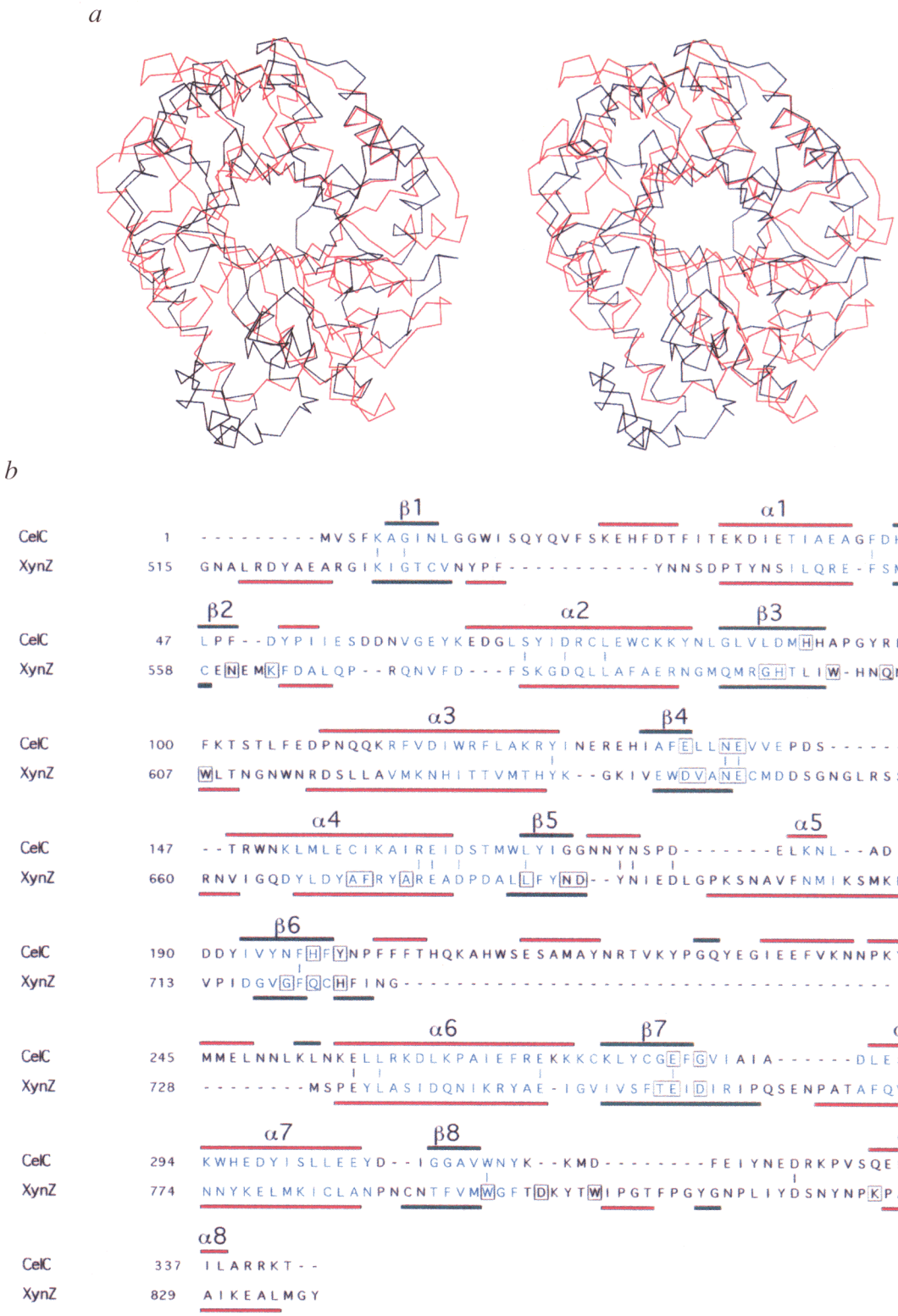

Fig. 3 *a*, Stereoview of the superimposed Cα backbones of CelC (in black) and XynZ (in red). *b*, Sequence alignment of the two proteins based on the three-dimensional structures. Sequence identities are indicated by vertical bars. Structurally equivalent residues are shown in blue (the r.m.s. deviation in Cα atoms is 1.9 Å for 159 equivalent positions). Helices (red bars) and β-strands (green bars) are indicated. Residues largely conserved within each family of enzymes are boxed. Family A cellulases have been classified in five subfamilies (A1–A5) on the basis of amino acid comparisons[6], with only a few critical residues strictly conserved across different classes. CelC belongs to subfamily A3.

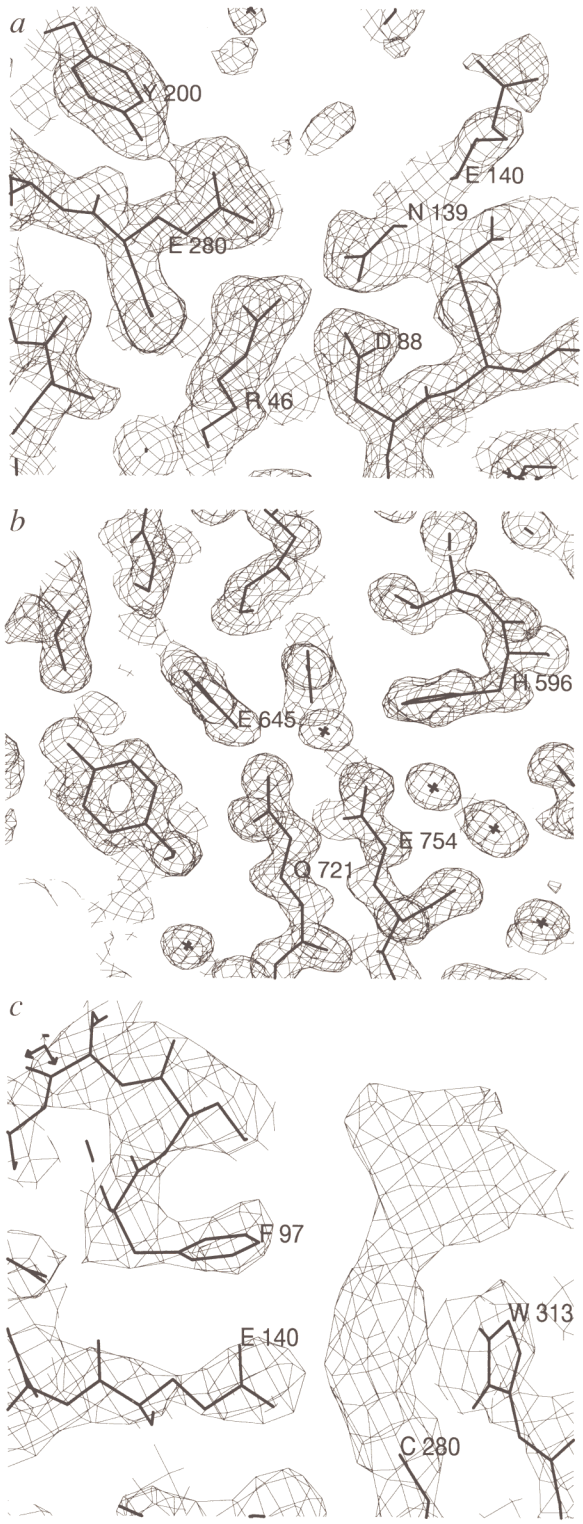

Fig. 4 Electron density of active-site residues. The maps (contoured at 1 σ level) were calculated with ($3F_o$-$2F_c$) amplitudes and phases calculated from the corresponding final models. *a*, Unliganded CelC in the orthorhombic space group at 2.15 Å resolution. Note the partially disordered side chain of Glu 140 pointing towards the solvent. *b*, Xylanase XynZ at 1.4 Å resolution. The peptide bond between His 596 and Thr 597 adopts a *cis* conformation. The catalytic residues Glu 645 and Glu 754 are indicated. *c*, $CelC_{E280C}$-cellotriose complex at 3 Å resolution. The conformation of Glu 140 and the path of the polypeptide chain for residues 95–103 are clearly visible in density. The unexplained electron density adjacent to position 280 is consistent with two bound glucosyl residues (not modelled).

residues Asn 92 and Glu 93 close to the centre of the active site. Glu 288 (but not Glu 93) has been proposed as the second catalytic residue by chemical labelling in barley glycanases[17]. The three-dimensional structures[16] are consistent with either residue Glu 288 or Glu 93 acting as the acid catalyst in the reaction[18] (Fig. 6), however the clear similarity of the active-site cleft with that of CelC (Fig. 5*a*) strongly suggests Glu 93 as the proton donor.

Several family A endoglucanases, particularly those of subfamily A4, display significant xylanase activity[19,20]. Conversely, family F xylanases are able to cleave chromogenic cellobiosides[21,22]. Family A enzymes, however, usually show a strong preference for β-1,4-glucan substrates, whereas family F enzymes predominantly hydrolyze xylan[23]. Some amino acid positions in the active site region, invariant within each family but differing between xylanases and cellulases, may be associated with the distinct specificity for cellulose or xylan substrates. This is the case for the two nucleophile-contacting residues (His 198 and Tyr 200 in CelC, Gln 721 and His 723 in XynZ), which are exposed to the binding cleft (Fig. 5). One difference involves Arg 46; its δ-guanido group interacts with several acidic and polar groups (Asn 9, Asp 88, Glu 136, Asn 139 and Glu 280) in CelC and is strictly conserved in family A cellulases (Fig. 3*b*). A smaller residue (valine or threonine) at the equivalent position in family F xylanases promotes the formation of a cavity (filled with water molecules in XynZ). This cavity may serve to accomodate side-chain substituents borne by natural xylans, which are characterized by a β-1,4-linked-D-xylopyranosyl backbone carrying a variable number of different monosaccharide or short oligosaccharide side chains.

Common evolutionary origin

A low level of amino acid sequence similarity and clear difference in substrate specificity led to the designation of two separate classes for glycanases of families A and F. Despite these differences, CelC (family A) and XynZ (family F) display similar structures and a common catalytic mechanism. Thus, the two enzyme families, as well as other β-glycanases presenting a similar overall fold and active site architecture[16], may derive from a common if distant ancestor whose divergent evolution has led to enzymes with different substrate specificities. Indeed, this hypothesis has been recently put forward by Jenkins and collaborators[18] and Henrissat and collaborators[24], who proposed that several established glycohydrolase families may actually form a superfamily of α/β barrel proteins having a similar disposition of catalytic residues.

Family B cellulases also share a similar (albeit irregular) α/β topology and have their active site at the same

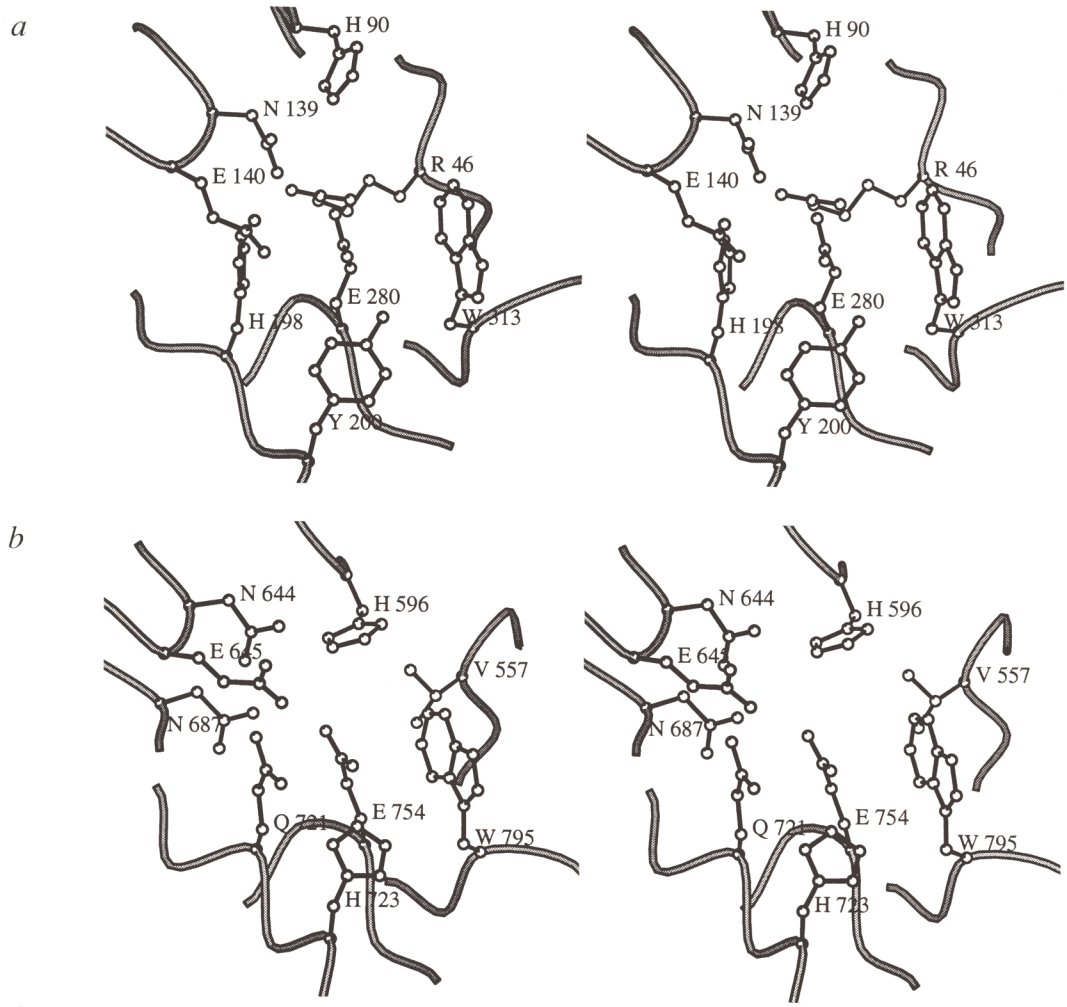

Fig. 5 Schematic stereoview of the active site in *a*, tetragonal CelC and *b*, XynZ showing conserved residues in family A cellulases and family F xylanases. A histidine residue in strand 3 (His 90 in CelC, His 596 in XynZ), which is strictly conserved in both families, albeit at non-equivalent positions in the sequence, occupies a similar position in space and is critical for enzymatic activity[7–9].

end of the barrel[25], yet catalysis proceeds by a different mechanism[4] (inversion of configuration) and the location of catalytic residues is quite different. Thus, any evolutionary relationship between cellulases of family B and enzymes of families A and F must be very distant. Glycosyl hydrolases of other families have similar acid/base mechanisms, but the three-dimensional structures now available reveal completely different protein folds[10,11,26,27]. Thus, the large diversity of enzymes required to metabolize the insoluble, recalcitrant polysaccharides of plant cell walls seems to have been generated both by adapting similar frameworks for different purposes and by recruiting altogether different framework structures.

Methods

Crystallization, structure determination and refinement of CelC. Endoglucanase CelC was expressed in *Escherichia coli*, purified, and crystallized as described[29]. Two crystal forms were obtained: orthorhombic ($P2_12_12_1$) with a=51.4 Å, b=84.3 Å, c=87.5 Å; and tetragonal ($P4_32_12$) with a=130.6 Å, c=69.6 Å. Native data from orthorhombic CelC crystals were collected using synchrotron radiation at EMBL-DESY, Hamburg (beamline X31, λ=0.9 Å). Data reduction (Table 1) was carried out with programs DENZO and SCALEPACK (Z. Otwinowski, unpublished program). Tetragonal crystals of CelC diffract to lower resolution; a preliminary data set was collected at 3.5 Å resolution using a Siemens/ Xentronics area detector mounted on a Rigaku rotating anode generator (28,416 observed reflections, 7,258 independent reflections, 99.8 % complete, R_{merge}=23.1 %).

The crystal structure of CelC was determined using MIR phases (Table 1) and density averaging between the two crystal forms. Calculations were performed with the CCP4 suite of programs[30] and the locally written program AVGMAP for density averaging (PMA, unpublished program). An electron density map of the orthorhombic crystal form, calculated with 3 Å MIR phases (figure of merit 0.77), further refined with the program DM from CCP4, revealed the molecular envelope and the general $(\alpha/\beta)_8$ topology of CelC, but was too discontinuous to allow polypeptide chain tracing.

The protein region of the MIR map (within the envelope corresponding to a single protein molecule) was then used as a search probe in rotation and translation function calculations, using a version of the program AMoRe (ref. 31) specifically modified for this purpose, to obtain the position of the enzyme and initial phase estimates in the tetragonal space group. Density averaging between the two crystal forms was then carried out with the program AVGMAP at 3.5 Å resolution. This procedure produced a continuous 3.5 Å map in both crystal forms, despite the high R_{merge} value of the tetragonal data set.

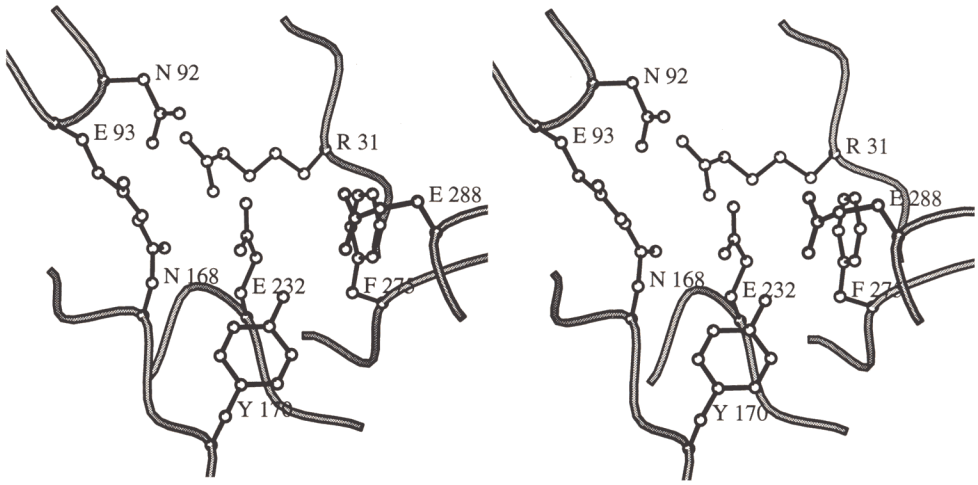

Fig. 6 Stereo diagram of the active site of barley (1-3,1-4)-β-glucanase[16] (PDB code 1GHR), representative of a distinct family of glycohydrolases (family 17 according to Henrisat[2]). Labelled residues are conserved within the family. Glu 232 is the nucleophile, whereas Glu 288 and Glu 93 are both favorably positioned to serve as the acid/base catalyst in the reaction (see text). The view is oriented as in Fig. 5.

Polypeptide chain tracing in the orthorhombic map was carried out with the program O[32]. Crystallographic refinement of the model at 2.15 Å was performed with XPLOR[33] using the parameter set proposed by Engh and Huber[34]. The final refinement parameters are listed in Table 1. Residues 95–103 are disordered and have not been modelled. All 312 non-glycine residues have ϕ,ψ values in allowed regions of the Ramachandran plot, according to PROCHECK[35].

Table 1 Data collection, phasing and refinement statistics

	Orthorhombic CelC			XynZ
Native data collection				
Resolution range (Å)		20–2.15		10–1.4
Observed reflections		160,050		343,930
Unique reflections		21,016		117,046
R_{merge} (%)[1]		8.6		8.3
Multiplicity		7.6		2.9
Completeness (%)		99.0		96.4
Isomorphous replacement				
Derivative	$HgCl_2$	MeHgCl	Me_3PbOAc	$HgCl_2$
Resolution (Å)	2.5	2.6	2.7	2.8
R_{merge} (%)	14.5	15.0	10.9	11.0
$R_{derivative}$ (%)[2]	26.4	24.8	20.6	28.0
Phasing power	1.6	1.4	1.0	2.7
Number of sites	4	4	1	14
Refinement				
Resolution range (Å)		6-2.15		10-1.4
Observed reflections		18,363		117,046
R_{factor} (%)[3]		16.9		18.2
Protein atoms		2,784		5,194
Water molecules		164		465
r.m.s.d. bond lengths (Å)		0.010		0.016
r.m.s.d. bond angles (degrees)		1.5		1.9
r.m.s.d. impropers (degrees)		1.4		1.7
r.m.s.d. B bonded atoms (Å^2)		1.5		1.7

[1] $R_{merge} = \sum_{hkl} |I - \langle I \rangle| / \sum_{hkl} \langle I \rangle$

[2] $R_{derivative} = \sum |\boldsymbol{F}_{der} - \boldsymbol{F}_{nat}| / \sum |\boldsymbol{F}_{nat}|$

[3] $R_{factor} = \sum |\boldsymbol{F}_{calc} - \boldsymbol{F}_{obs}| / \sum |\boldsymbol{F}_{obs}|$

Crystallization, structure determination and refinement of XynZ. Purification, and crystallization of the catalytic domain of XynZ (residues 508–837) have been described[22,28]. Crystals are triclinic, space group P1, with a=47.10 Å, b=51.10 Å, c=70.74 Å, α = 100.54°, β = 83.79° and γ = 101.64°, with two molecules in the unit cell. A native data set was collected using synchrotron radiation (Table 1) as described above for CelC. The structure of XynZ was determined at 2.8 Å resolution using single isomorphous replacement phases (Table 1) and density averaging between the two molecules in the asymmetric unit. The positions of eight major Hg binding sites were found by direct methods using the program SHELX-90 (ref. 36), and six additional minor sites were identified by analysis of residual Fourier synthesis. Heavy atom sites were distributed in two spatial clusters related to each other by a non-crystallographic symmetry operator which corresponded to the highest peak in a self-rotation function map. An initial map calculated at 2.8 Å resolution with solvent-flattened phases revealed the molecular envelopes, but exhibited poor connectivity. This map was significantly improved by density averaging between the two molecules in the asymmetric unit using the program AVGMAP, and an initial model could be built as a discontinuous polypeptide (chain tracing was facilitated by the recognition of a typical $(\alpha/\beta)_8$ barrel topology at an early stage in map interpretation).

After obtaining a complete model at 2.8 Å resolution, the structure was refined with program ARP (ref. 37), first at 1.8 Å and then including all data between 10 and 1.4 Å resolution. During this procedure, minor modifications were introduced by inspection of $(3\boldsymbol{F}_o\text{-}2\boldsymbol{F}_c)$ electron density maps. A close analysis of the water model introduced by ARP was carried out to exclude solvent sites having poor density or accounting for alternative side-chain conformations. The revised model was subjected to a final round of simulated annealing refinement with X-PLOR using all diffraction intensities. The final atomic model (Fig. 4*b*) has good overall stereochemistry (Table 1); among the non-glycine and non-proline residues, 517 (91 %) have ϕ,ψ values in the 'most favoured' regions[35] of the Ramachandran plot. The r.m.s. deviation in main chain atom positions for the two molecules is 0.34 Å.

The atomic coordinates of orthorhombic CelC and XynZ are being deposited with the Protein Data Bank, Chemistry Department, Broohaven National Laboratory, Upton, NY 11973.

Structure determination of the CelC$_{E280C}$-cellotriose complex. Tetragonal crystals of a complex between cellotriose and an inactive mutant of CelC in which the nucleophilic glutamate was substituted by a cysteine residue (CelC$_{E280C}$) were obtained as for wild-type CelC[29]. Diffraction data was collected to 3 Å using a Siemens/Xentronics area detector and a rotating anode generator. A total of 27,255 observed reflections was reduced to 11,641 unique reflections (93.3 % complete to 3 Å) with an R_{merge} of 15 %. The refined model of orthorhombic CelC was placed in the tetragonal unit cell and refined with X-PLOR at 3 Å resolution as described above. Inspection of ($2\boldsymbol{F}_o$-$\boldsymbol{F}_c$) and omit maps in the initial refinement cycles allowed unambiguous tracing of the loop 95–103 (Fig. 4*c*), which was disordered in the orthorhombic form. Electron density consistent with the bound substrate was observed (Fig. 4*c*) but not modelled. After refinement of individual temperature factors, the crystallographic *R*-factor is 18.6 % for 10,208 reflections in the 8–3 Å resolution range for a model consisting of protein atoms only. The r.m.s. deviations of bond lengths, bond angles and improper angles from ideal values are 0.011 Å, 1.75° and 1,55°, respectively. The r.m.s. deviation in temperature factors of bonded atoms is 1.7 Å^2. All 304 non-glycine and non-proline residues have ϕ,ψ values in allowed regions of the Ramachandran plot.

Least-squares superposition of CelC and XynZ. Superposition of atomic coordinates was carried out iteratively using the program SUPERPK (PMA, unpublished program). During each cycle, optimal least-squares superposition of Cα coordinates from a set of pairwise-equivalent residues was performed as described by Kabsch[38], and the equivalence set was then updated to include new (or exclude old) equivalent positions according to distance criteria. The iteration typically converged to an invariant equivalence set in a few cycles. A distance cutoff of 3.2 Å and equivalent zones of three or more consecutive residues were used to calculate the structural alignment presented in Fig. 3*b*.

Received 3 March; accepted 16 May 1995

1. Gilkes, N.R., Henrissat, B., Kilburn, D.G., Miller, R.C., Jr & Warren, R.A.J. Domains in microbial β-1,4-glycanases: Sequence conservation, function, and enzyme families. *Microbiol. Rev.* **55**, 303–315 (1991).
2. Henrissat, B. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **280**, 309–316 (1991).
3. Henrissat, B. & Bairoch, A. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **293**, 781–788 (1993).
4. Gebler, J. *et al.* Stereoselective hydrolysis catalyzed by related β-1,4-glucanases and β-1,4-xylanases. *J. biol. Chem.* **267**, 12559–12561 (1992).
5. Banner, D.W. *et al.* Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5 Å resolution using amino acid sequence data. *Nature* **255**, 609–614 (1975).
6. Wang, Q. *et al.* Glu 280 is the nucleophile in the active site of *Clostridium thermocellum* CelC, a family A endo-β-1,4-glucanase. *J. biol. Chem.* **268**, 14096–14102 (1993).
7. Navas, J. & Béguin, P. Site-directed mutagenesis of conserved residues in *Clostridium thermocellum* endoglucanase CelC. *Biochem. biophys. res. Comm.* **189**, 807–812 (1992).
8. Py, B., Bortoli-German, I., Haiech, J., Chippaux, M. & Barras, F. Cellulase EGZ of *Erwinia chrysanthemi*: structural organization and importance of His 98 and Glu 133 residues for catalysis. *Prot. Engng* **4**, 325–333 (1991).
9. Belaich, A. *et al.* The catalytic domain of endoglucanase A from *Clostridium cellulolyticum* - Effects of arginine-79 and histidine-122 mutations on catalysis. *J. Bacteriol.* **174**, 4677–4682 (1992).
10. Davies, G.J. *et al.* Structure and function of endoglucanase V. *Nature* **365**, 362–364 (1993).
11. Törrönen, A., Harkki, A. & Rouvinen, J. Three-dimensional structure of endo-1,4-β-xylanase II from *Trichoderma reesei*: Two conformational states in the active site. *EMBO J.* **13**, 2493–2501 (1994).
12. Derewenda, U. *et al.* Crystal structure, at 2.6 Å resolution, of the *Streptomyces lividans* xylanase A, a member of the F family of β-1,4-D-glycanases. *J. biol. Chem.* **269**, 20811–20814 (1994).
13. White, A., Withers, S.G., Gilkes, N.R. & Rose, D.R. Crystal structure of the catalytic domain of the β-1,4-glycanase Cex from *Cellulomonas fimi. Biochemistry* **33**, 12546–12552 (1994).
14. Harris, G.W. *et al.* Structure of the catalytic core of the family F xylanase from *Pseudomonas fluorescens* and identification of the xylopentaose-binding sites. *Structure* **2**, 1107–1116 (1994).
15. Tull, D., Withers, S.G., Gilkes, N.R., Kilburn, D.G., Warren, R.A.J. & Aebersold, R. Glutamic acid 274 is the nucleophile in the active site of a "retaining" exoglucanase from *Cellulomonas fimi. J. biol. Chem.* **266**, 15621–15625 (1991).
16. Varghese, J.N., Garrett, T.P.J., Colman, P.M., Chen, L., Høj, P.B. & Fincher, G.B. Three-dimensional structure of two plant beta-glucan endohydrolases with distinct substrate specificities. *Proc. natn. Acad. Sci. U.S.A.* **91**, 2785–2789 (1994).
17. Chen, L., Fincher, G.B. & Høj, P.B. Evolution of polysaccharide hydrolase substrate specificity. *J. biol. Chem.* **268**, 13318–13326 (1993).
18. Jenkins, J., Lo Leggio, L., Harris, G. & Pickersgill, R. β-glucosidase, β-galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily of enzymes with 8-fold β/α architecture and with two conserved glutamates near the carboxy-terminal ends of β-strands four and seven. *FEBS Let.* **in the press**.
19. Fierobe, H.-P. *et al.* Characterization of endoglucanase A from *Clostridium cellulolyticum. J. Bacteriol.* **173**, 7956–7962 (1991).
20. Yagüe, E., Béguin, P. & Aubert, J.-P. Nucleotide sequence and deletion analysis of the cellulase-encoding gene *celH* of *Clostridium thermocellum. Gene* **89**, 61–67 (1990).
21. Gilkes, N.R., Langsford, M. L., Kilburn, D. G., Miller Jr, R. C. & Warren, R. A. J. Mode of action and substrate specificities of cellulases from cloned bacterial genes. *J. biol. Chem.* **259**, 10455–10459 (1984).
22. Grépinet, O., Chebrou, M.-C. & Béguin, P. Nucleotide sequence and deletion analysis of the xylanase gene (*xynZ*) of *Clostridium thermocellum. J. Bacteriol.* **170**, 4582–4588 (1988).
23. Claeyssens, M. & Henrissat, B. Specificity mapping of cellulolytic enzymes - classification into families of structurally related proteins confirmed by biochemical analysis. *Prot. Sci.* **1**, 1293–1297 (1992).
24. Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.-P. & Davies, G. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc. natn. Acad. Sci. U.S.A.*, **in the press**.
25. Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J.K.C. & Jones, T.A. Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei. Science* **249**, 380–386 (1990).
26. Divne, C. *et al.* The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei. Science* **265**, 524–528 (1994)
27. Juy, M. *et al.* Crystal structure of a thermostable bacterial cellulose-degrading enzyme. *Nature* **357**, 89–91 (1992).
28. Souchon, H., Spinelli, S., Béguin, P. & Alzari, P.M. Crystallization and preliminary diffraction analysis of the catalytic domain of xylanase Z from *Clostridium thermocellum. J. molec. Biol.* **235**, 1348–1350 (1994).
29. Dominguez, R., Souchon, H. & Alzari, P.M. Characterization of two crystal forms of *Clostridium thermocellum* endoglucanase CelC. *Proteins* **19**, 158–160 (1994).
30. CCP4, The SERC (UK) Collaborative Computing Project No. 4: *A Suite of Programs for Protein Crystallography* (Daresbury, UK; 1979).
31. Navaza, J. AMoRe: an automated package for molecular replacement. *Acta crystallogr.* **A50**, 157–163 (1994).
32. Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta crystallogr.* **A47**, 110–119 (1991).
33. Brünger, A.T., Kuriyan, J. & Karplus, M. Crystallographic R-factor refinement by molecular dynamics. *Science* **235**, 458–460 (1987).
34. Engh, R.A. & Huber, R. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta crystallogr.* **A47**, 392–400 (1991).
35. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. appl. Crystallogr.* **26**, 283–291 (1993).
36. Sheldrick, G.M. Phase annealing in SHELX-90: direct methods for larger structures. *Acta crystallogr.* **A46**, 467–473 (1990).
37. Lamzin, V.S. & Wilson, K.S. Automated refinement of protein models. *Acta crystallogr.* **D49**, 129–147 (1993).
38. Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta crystallogr.* **A32**, 922–923 (1976).
39. Kraulis, P.J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. appl. Crystallogr.* **24**, 946–950 (1991).